

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

日本国特許庁

PATENT OFFICE
JAPANESE GOVERNMENT

03.03.00

REC'D 17 MAR 2000

WIPO PCT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日

Date of Application:

1999年 3月 4日

出願番号

Application Number:

平成11年特許願第057467号

出願人

Applicant (s):

ソニー株式会社

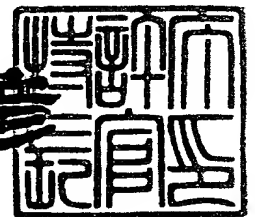
PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2000年 2月 4日

特許庁長官
Commissioner,
Patent Office

近藤 隆彦



出証番号 出証特2000-3003277

【書類名】 特許願

【整理番号】 9801154103

【提出日】 平成11年 3月 4日

【あて先】 特許庁長官殿

【国際特許分類】 G06K 9/62
G10L 3/00

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社
内

【氏名】 包 洪長

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100082131

【弁理士】

【氏名又は名称】 稲本 義雄

【電話番号】 03-3369-6479

【手数料の表示】

【予納台帳番号】 032089

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9708842

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 パターン認識装置および方法、並びに提供媒体

【特許請求の範囲】

【請求項 1】 入力されるデータの特徴分布を、所定数のモデルのうちのいずれかに分類するパターン認識装置において、

前記入力されるデータのパターンを特徴分布として抽出する抽出手段と、

前記所定数のモデルを記憶する記憶手段と、

前記抽出手段が抽出した特徴分布を、前記所定数のモデルのうちのいずれかに分類する分類手段と、

前記データが入力される直前に入力されたノイズに基づいて、前記データが存在しない状態に対応する前記モデルを生成し、前記記憶手段に記憶されている対応するものを更新する生成手段と

を含むことを特徴とするパターン認識装置。

【請求項 2】 前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前期生成手段は、前記データが存在しない状態に対応する前記モデルの期待値を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する期待値の平均として生成し、前記データが存在しない状態に対応する前記モデルの分散を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する分散の平均として生成する

ことを特徴とする請求項 1 に記載のパターン認識装置。

【請求項 3】 前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前期生成手段は、前記データが存在しない状態に対応する前記モデルの期待値および分散を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する期待値の平均を用いて生成する

ことを特徴とする請求項 1 に記載のパターン認識装置。

【請求項 4】 前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前

期生成手段は、前記データが存在しない状態に対応する前記モデルの確率分布を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する統計量の線形結合に基づいて生成する

ことを特徴とする請求項1に記載のパターン認識装置。

【請求項5】 前記データが存在しない状態の特徴分布、および、前記データが存在しない状態に対応する前記モデルの確率分布が正規分布である場合、前期生成手段は、前記データが存在しない状態に対応する前記モデルの確率分布を、前記データが存在しない状態の特徴分布の各コンポーネントに対応する統計母集団の和に基づいて生成する

ことを特徴とする請求項1に記載のパターン認識装置。

【請求項6】 入力されるデータの特徴分布を、所定数のモデルのうちのいずれかに分類するパターン認識装置のパターン認識方法において、

前記入力されるデータのパターンを特徴分布として抽出する抽出ステップと、

前記所定数のモデルを記憶する記憶ステップと、

前記抽出ステップで抽出した特徴分布を、前記所定数のモデルのうちのいずれかに分類する分類ステップと、

前記データが入力される直前に入力されたノイズに基づいて、前記データが存在しない状態に対応する前記モデルを生成し、前記記憶ステップで記憶された対応するものを更新する生成ステップと

を含むことを特徴とするパターン認識方法。

【請求項7】 入力されるデータの特徴分布を、所定数のモデルのうちのいずれかに分類するパターン認識装置に、

前記入力されるデータのパターンを特徴分布として抽出する抽出ステップと、

前記所定数のモデルを記憶する記憶ステップと、

前記抽出ステップで抽出した特徴分布を、前記所定数のモデルのうちのいずれかに分類する分類ステップと、

前記データが入力される直前に入力されたノイズに基づいて、前記データが存在しない状態に対応する前記モデルを生成し、前記記憶ステップで記憶された対応するものを更新する生成ステップと

を含む処理を実行させるコンピュータが読み取り可能なプログラムを提供することを特徴とする提供媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、パターン認識装置および方法、並びに提供媒体に関し、特に、ノイズ環境下において発話された単語を認識するパターン認識装置および方法、並びに提供媒体に関する。

【0002】

【従来の技術】

従来より、ノイズ環境下において発話された単語を識別する方法が考案されており、その代表的な方法としては、PMC(Parallel Model Combination)法、SS/NS S(Spectral Subtraction/Nonlinear Spectral Subtraction)法、SFE(Stochastic Feature Extraction)法等が知られている。

【0003】

上述したいずれの方法においても、発話音声に環境ノイズが混在している音声データの特徴量が抽出され、その特徴量が、予め登録されている複数の単語に対応する音響モデルのうちのいずれに最も適合するかが判定されて、最も適合する音響モデルに対応する単語が認識結果として出力される。

【0004】

上述した方法の特徴は、以下の通りである。すなわち、PMC法は、環境ノイズの情報を直接的に音響モデルに取り込んでいるので認識性能は良いが、計算コストが高い（高度な演算を必要とするので、装置の規模が大型化する、処理に要する時間が長い等）。SS/NSS法は、音声データの特徴量を抽出する段階において、環境ノイズを除去している。したがって、PMC法よりも計算コストが低く、現在、多く用いられている方法である。なお、SS/NSS法は、音声データの特徴量をベクトルとして抽出する。SFE法は、SS/NSS法と同様に、ミックス信号の特徴量を抽出する段階において、環境ノイズを除去するが、特徴量を確率分布として抽出する。

【0 0 0 5】

【発明が解決しようとする課題】

ところで、SFE法においては、音声認識の段階で環境ノイズが直接的に反映されていない、すなわち、環境ノイズの情報が直接的に無音音響モデルに取り込まれていないので、認識性能が劣る課題があった。

【0 0 0 6】

また、環境ノイズの情報が直接的に無音音響モデルに取り込まれていないことに起因して、音声認識が開始された時点から発話が始まるまでの時間が長くなるにつれて認識性能が低下する課題があった。

【0 0 0 7】

本発明はこのような状況に鑑みてなされたものであり、環境ノイズの情報を用いて無音音響モデルを補正することにより、音声認識が開始された時から発話が始まるまでの時間が長くなるに伴って認識性能が低下することを抑止するようにするものである。

【0 0 0 8】

【課題を解決するための手段】

請求項 1 に記載のパターン認識装置は、入力されるデータのパターンを特徴分布として抽出する抽出手段と、所定数のモデルを記憶する記憶手段と、抽出手段が抽出した特徴分布を、所定数のモデルのうちのいずれかに分類する分類手段と、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶手段に記憶されている対応するものを更新する生成手段とを含むことを特徴とする。

【0 0 0 9】

請求項 6 に記載のパターン認識方法は、入力されるデータのパターンを特徴分布として抽出する抽出ステップと、所定数のモデルを記憶する記憶ステップと、抽出ステップで抽出した特徴分布を、所定数のモデルのうちのいずれかに分類する分類ステップと、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶ステップで記憶された対応するものを更新する生成ステップとを含むことを特徴とする。

【0 0 1 0】

請求項 7 に記載の提供媒体は、入力されるデータのパターンを特徴分布として抽出する抽出ステップと、所定数のモデルを記憶する記憶ステップと、抽出ステップで抽出した特徴分布を、所定数のモデルのうちのいずれかに分類する分類ステップと、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶ステップで記憶された対応するものを更新する生成ステップとを含む処理をパターン認識装置に実行させるコンピュータが読み取り可能なプログラムを提供することを特徴とする。

【0 0 1 1】

請求項 1 に記載のパターン認識装置、請求項 6 に記載のパターン認識方法、および請求項 7 に記載の提供媒体においては、入力されるデータのパターンが特徴分布として抽出され、所定数のモデルが記憶されて、抽出された特徴分布が、所定数のモデルのうちのいずれかに分類される。また、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルが生成されて記憶されている対応するものが更新される。

【0 0 1 2】

【発明の実施の形態】

本発明を適用した音声認識装置の構成例について、図 1 を参照して説明する。この音声認識装置において、マイクロフォン 1 は、認識対象である発話音声、環境ノイズとともに集音し、フレーム化部 2 に出力する。フレーム化部 2 は、マイクロフォン 1 から入力される音声データを、所定の時間間隔（例えば、10ms）で取り出し、その取り出したデータを、1 フレームのデータとして出力する。フレーム化部 2 が出力する 1 フレーム単位の音声データは、そのフレームを構成する時系列の音声データそれぞれをコンポーネントとする観測ベクトル a として、ノイズ観測区間抽出部 3、および特徴抽出部 5 に供給される。

【0 0 1 3】

ここで、以下、適宜、第 t フレームの音声データである観測ベクトルを、 $a(t)$ と表す。

【0014】

ノイズ観測区間抽出部3は、フレーム化部2から入力されるフレーム化された音声データを所定の時間（Mフレーム分以上）だけバッファリングしており、図2に示すように、発話スイッチ4がオンとされたタイミング t_b からMフレーム分だけ以前のタイミング t_a までをノイズ観測区間 T_n とし、ノイズ観測区間 T_n におけるMフレーム分の観測ベクトル a を抽出して特徴抽出部5、および無音音響モデル補正部7に出力する。

【0015】

発話スイッチ4は、ユーザが発話を開始するときにユーザ自身がオンとするスイッチである。したがって、発話スイッチ4がオンとされたタイミング t_b 以前（ノイズ観測区間 T_n ）の音声データには、発話音声は含まれず、環境ノイズだけが存在する。また、発話スイッチ4がオンとされたタイミング t_b から発話スイッチ4がオフとされるタイミング t_d までは、音声認識区間とされて、その区間の音声データが音声認識の対象とされる。

【0016】

特徴抽出部5は、ノイズ観測区間抽出部3から入力されるノイズ観測区間 T_n の環境ノイズだけが存在する音声データに基づいて、フレーム化部2から入力される、タイミング t_b 以降の音声認識区間の観測ベクトル a から環境ノイズ成分を除去して、その特徴量を抽出する。すなわち、特徴抽出部5は、例えば、観測ベクトル a としての真（環境ノイズが除去された）の音声データをフーリエ変換し、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトル y を算出する。なお、パワースペクトラムの算出方法は、フーリエ変換によるものに限定されるものではない。すなわち、パワースペクトラムは、その他、例えば、いわゆるフィルタバンク法などによって求めることも可能である。

【0017】

さらに、特徴抽出部5は、観測ベクトル a としての音声データに含まれる音声を、その特徴量の空間（特徴ベクトル空間）に写像したときに得られる、その特徴ベクトル空間上の分布を表すパラメータ（以下、特徴分布パラメータと記述す

る) z を、算出した特徴ベクトル y に基づいて算出し、音声認識部 6 に供給する。

【0018】

図 3 は、図 1 の特徴抽出部 5 の詳細な構成例を示している。フレーム化部 2 から入力される観測ベクトル a は、特徴抽出部 5 において、パワースペクトラム分析部 11 に供給される。パワースペクトラム分析部 11 では、観測ベクトル a が、例えば、FFT（高速フーリエ変換）アルゴリズムによってフーリエ変換され、これにより、音声の特徴量であるパワースペクトラムが、特徴ベクトルとして抽出される。なお、ここでは、1 フレームの音声データとしての観測ベクトル a が、 D 個のコンポーネントからなる特徴ベクトル（ D 次元の特徴ベクトル）に変換されるものとする。

【0019】

ここで、第 t フレームの観測ベクトル $a(t)$ から得られる特徴ベクトルを $y(t)$ と表す。また、特徴ベクトル $y(t)$ のうち、真の音声のスペクトル成分を $x(t)$ と、環境ノイズのスペクトル成分を $u(t)$ と表す。この場合、真の音声のスペクトル成分 $x(t)$ は、次式 (1) で表される。

【数 1】

$$x(t) = y(t) - u(t) \quad \dots (1)$$

ただし、ここでは、環境ノイズが不規則な特性を有し、また、観測ベクトル $a(t)$ としての音声データは、真の音声成分に環境ノイズを加算したものであると仮定されている。

【0020】

一方、ノイズ観測区間抽出部 3 から入力される音声データ（環境ノイズ）は、特徴検出部 5 において、ノイズ特性算出部 13 に入力される。ノイズ特性算出部 13 では、ノイズ観測区間 T_n における環境ノイズの特性が求められる。

【0021】

すなわち、ここでは、音声認識区間における環境ノイズのパワースペクトラム $u(t)$ の分布が、その音声認識区間の直前のノイズ観測区間 T_n における環境

ノイズと同一であるとし、さらに、その分布が正規分布であるとして、ノイズ特性算出部 13 において、環境ノイズの平均値（平均ベクトル）と分散（分散マトリクス）が求められる。

【0022】

なお、平均ベクトル μ' と分散マトリクス Σ' は、次式 (2), (3) にしたがって求めることができる。

【数 2】

$$\begin{aligned}\mu' (i) &= \frac{1}{M} \sum_{t=1}^M y(t)(i) \\ \Sigma' (i,j) &= \frac{1}{M} \sum_{t=1}^M (y(t)(i) - \mu' (i))(y(t)(j) - \mu' (j)) \\ &\quad \dots (2)\end{aligned}$$

ただし、 $\mu' (i)$ は、平均ベクトル μ' の i 番目のコンポーネントを表す ($i = 1, 2, \dots, D$)。また、 $y(t)(i)$ は、第 t フレームの特徴ベクトルの i 番目のコンポーネントを表す。さらに、 $\Sigma' (i, j)$ は、分散マトリクス Σ' の、第 i 行、第 j 列のコンポーネントを表す ($j = 1, 2, \dots, D$)。

【0023】

ここで、計算量の低減のために、環境ノイズについては、特徴ベクトル y の各コンポーネントが、互いに無相関であると仮定する。この場合、次式に示すように、分散マトリクス Σ' は、対角成分以外は 0 となる。

【数 3】

$$\Sigma' (i,j) = 0, i \neq j \quad \dots (3)$$

【0024】

ノイズ特性算出部 13 では、以上のようにして、環境ノイズの特性としての平均ベクトル μ' および平均値 Σ' が求められ、特徴分布パラメータ算出部 12 に供給される。

【 0 0 2 5 】

一方、パワースペクトラム分析部 1 1 の出力、すなわち、環境ノイズを含む発話音声の特徴ベクトル y は、特徴分布パラメータ算出部 1 2 に供給される。特徴分布パラメータ算出部 1 2 では、パワースペクトラム分析部 1 1 からの特徴ベクトル y 、およびノイズ特性算出部 1 3 からの環境ノイズの特性に基づいて、真の音声のパワースペクトラムの分布（推定値の分布）を表す特徴分布パラメータが算出される。

【 0 0 2 6 】

すなわち、特徴分布パラメータ算出部 1 2 では、真の音声のパワースペクトラムの分布が正規分布であるとして、その平均ベクトル ξ と分散マトリクス Ψ が、特徴分布パラメータとして、次式（4）乃至（7）にしたがって計算される。

【数 4】

$$\begin{aligned}
 \xi(t)(i) &= E[x(t)(i)] \\
 &= E[y(t)(i) - u(t)(i)] \\
 &= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i)) \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i) \\
 &= \frac{y(t)(i) \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) - \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \\
 &= y(t)(i) - \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \quad \dots (4)
 \end{aligned}$$

【数 5】

$i=j$ のとき

$$\begin{aligned}\Psi(t)(i,j) &= V[x(t)(i)] \\ &= E[(x(t)(i))^2] - (E[x(t)(i)])^2 \\ &= E[(x(t)(i))^2] - (\xi(t)(i))^2\end{aligned}$$

$i \neq j$ のとき

$$\Psi(t)(i,j) = 0 \quad \dots (5)$$

【数 6】

$$E[(x(t)(i))^2] = E[(y(t)(i) - u(t)(i))^2]$$

$$\begin{aligned}&= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i))^2 \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i) \\&= \frac{1}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \times \left\{ (y(t)(i))^2 \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) \right. \\&\quad - 2y(t)(i) \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i) \\&\quad \left. + \int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i) \right\} \\&= (y(t)(i))^2 - 2y(t)(i) \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \\&\quad + \frac{\int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \quad \dots (6)\end{aligned}$$

【数 7】

$$P(u(t)(i)) = \frac{1}{\sqrt{2\pi \Sigma'(i,i)}} e^{-\frac{1}{2\Sigma'(i,i)} (u(t)(i) - \mu'(i))^2} \dots (7)$$

【0027】

ここで、 $\xi(t)(i)$ は、第 t フレームにおける平均ベクトル $\xi(t)$ の i 番目のコンポーネントを表す。また、 $E[\]$ は、 $[\]$ 内の平均値を意味する。 $x(t)(i)$ は、第 t フレームにおける真の音声のパワースペクトラム $x(t)$ の i 番目のコンポーネントを表す。さらに、 $u(t)(i)$ は、第 t フレームにおける環境ノイズのパワースペクトラムの i 番目のコンポーネントを表し、 $P(u(t)(i))$ は、第 t フレームにおける環境ノイズのパワースペクトラムの i 番目のコンポーネントが $u(t)(i)$ である確率を表す。ここでは、環境ノイズの分布として正規分布を仮定しているので、 $P(u(t)(i))$ は、式 (7) に示したように表される。

【0028】

また、 $\Psi(t)(i, j)$ は、第 t フレームにおける分散 $\Psi(t)$ の、第 i 行、第 j 列のコンポーネントを表す。さらに、 $V[\]$ は、 $[\]$ 内の分散を表す。

【0029】

特徴分布パラメータ算出部 12 では、以上のようにして、各フレームごとに、平均ベクトル ξ および分散マトリクス Ψ が、真の音声の特徴ベクトル空間上での分布（ここでは、真の音声の特徴ベクトル空間上での分布が正規分布であると仮定した場合の、その分布）を表す特徴分布パラメータとして求められる。

【0030】

その後、音声認識区間の各フレームにおいて求めた特徴分布パラメータは、音声認識部 6 に出力される。すなわち、いま、音声認識区間が T フレームであったとし、その T フレームそれぞれにおいて求められた特徴分布パラメータを、 $z(t) = \{\xi(t), \Psi(t)\}$ ($t = 1, 2, \dots, T$) と表すと、特徴分布パラメータ算出部 12 は、特徴分布パラメータ（系列） $Z = \{z(1), z(2$

), ..., z(T)} を、音声認識部 6 に供給する。

【0031】

図 1 に戻る。音声認識部 6 は、特徴抽出部 5 から入力される特徴分布パラメータ Z を、所定数 K の音響モデルと 1 個の無音音響モデルのうちのいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。すなわち、音声認識部 6 は、例えば、無音区間に対応する識別関数（特徴パラメータ Z が無音音響モデルに分類されるかを識別するための関数）と、所定数 K の単語それぞれに対応する識別関数（特徴パラメータ Z がいずれの音響モデルに分類されるかを識別するための関数）とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部 5 からの特徴分布パラメータ Z を引数として計算する。そして、その関数値が最大である音響モデル（単語、または無音区間）が認識結果として出力される。

【0032】

図 4 は、図 1 の音声認識部 6 の詳細な構成例を示している。特徴抽出部 5 の特徴分布パラメータ算出部 12 から入力される特徴分布パラメータ Z は、識別関数演算部 21-1 乃至 21- k 、および識別関数演算部 21- s 、に供給される。識別関数演算部 21- k ($k=1, 2, \dots, K$) は、 K 個の音響モデルのうちの k 番目に対応する単語を識別するための識別関数 $G_k(Z)$ を記憶しており、特徴抽出部 5 からの特徴分布パラメータ Z を引数として、識別関数 $G_k(Z)$ を演算する。識別関数演算部 21- s は、無音音響モデルに対応する無音区間を識別するための識別関数 $G_s(Z)$ を記憶しており、特徴抽出部 5 からの特徴分布パラメータ Z を引数として、識別関数 $G_s(Z)$ を演算する。

【0033】

なお、音声認識部 6 では、例えば、HMM(Hidden Markov Model)法を用いて、クラスとしての単語または無音区間の識別（認識）が行われる。

【0034】

HMM法について、図 5 を参照して説明する。同図において、HMMは、 H 個の状態 q_1 乃至 q_H を有しており、状態の遷移は、自身への遷移と、右隣の状態への遷移のみが許されている。また、初期状態は、最も左の状態 q_1 とされ、最終状態は

、最も右の状態 q_H とされており、最終状態 q_H からの状態遷移は禁止されている。このように、自身よりも左にある状態への遷移のないモデルは、left-to-right モデルと呼ばれ、音声認識では、一般に、left-to-right モデルが用いられる。

【0035】

いま、HMMの k クラスを識別するためのモデルを、 k クラスモデルというすると、 k クラスモデルは、例えば、最初に状態 q_h にいる確率（初期状態確率） $\pi_k(q_h)$ 、ある時刻（フレーム） t において、状態 q_i にいて、次の時刻 $t+1$ において、状態 q_j に状態遷移する確率（遷移確率） $a_k(q_i, q_j)$ 、および状態 q_i から状態遷移が生じるときに、その状態 q_i が、特徴ベクトル O を出力する確率（出力確率） $b_k(q_i)(O)$ によって規定される（ $h=1, 2, \dots, H$ ）。

【0036】

そして、ある特徴ベクトル系列 O_1, O_2, \dots が与えられた場合、例えば、そのような特徴ベクトル系列が観測される確率（観測確率）が最も高いモデルのクラスが、その特徴ベクトル系列の認識結果とされる。

【0037】

ここでは、この観測確率が、識別関数 $G_k(Z)$ によって求められる。すなわち、識別関数 $G_k(Z)$ は、特徴分布パラメータ（系列） $Z = \{z_1, z_2, \dots, z_T\}$ に対する最適状態系列（最適な状態の遷移のしていき方）において、そのような特徴分布パラメータ（系列） $Z = \{z_1, z_2, \dots, z_T\}$ が観測される確率を求めるものとして、次式（8）で与えられる。

【数8】

$$g_k(Z) = \max_{q_1, q_2, \dots, q_T} \pi_k(q_1) \cdot b'_k(q_1)(z_1) \cdot a_k(q_1, q_2) \cdot b'_k(q_2)(z_2) \\ \cdot \dots \cdot a_k(q_{T-1}, q_T) \cdot b'_k(q_T)(z_T) \quad \dots (8)$$

【0038】

ここで、 $b'_k(q_i)(z_j)$ は、出力が z_j で表される分布であるときの出力確率を表す。状態遷移時に各特徴ベクトルを出力する確率である出力確率 b_k （

$s)$ (O_t) には、ここでは、例えば、特徴ベクトル空間上のコンポーネントに
 相関がないものとして、正規分布関数が用いられている。この場合、入力が z_t
 で表される分布であるとき、出力確率 $b_k'(s)(z_t)$ は、平均ベクトル μ_k
 (s) と分散マトリクス $\Sigma_k(s)$ とによって規定される確率密度関数 $P_k^m(s)$
 (x) 、および第 t フレームの特徴ベクトル (ここでは、パワースペクトラム
 $) x$ の分布を表す確率密度関数 $P^f(t)(x)$ を用いて、次式 (9) により求
 めることができる。

【数 9】

$$\begin{aligned} b_k'(s)(z_t) &= \int P^f(t)(x) P_k^m(s)(x) dx \\ &= \prod_{i=1}^D P(s)(i)(\xi(t)(i), \Psi(t)(i, i)) \\ &\quad k=1, 2, \dots, K : s=q_1, q_2, \dots, q_T : T=1, 2, \dots, T \\ &\quad \dots (9) \end{aligned}$$

ただし、式 (9) における積分の積分区間は、 D 次元の特徴ベクトル空間 (ここ
 では、パワースペクトラム空間) の全体である。

【0039】

また、式 (9) において、 $P(s)(i)(\xi(t)(i), \Psi(t)(i, i))$ は、次式 (10) で表される。

【数 10】

$$\begin{aligned} &P(s)(i)(\xi(t)(i), \Psi(t)(i, i)) \\ &= \frac{1}{\sqrt{2\pi(\Sigma_k(s)(i, i) + \Psi(t)(i, i))}} e^{-\frac{(\mu_k(s)(i) - \xi(t)(i))^2}{2(\Sigma_k(s)(i, i) + \Psi(t)(i, i))}} \\ &\quad \dots (10) \end{aligned}$$

ただし、 $\mu_k(s)(i)$ は、平均ベクトル $\mu_k(s)$ の i 番目のコンポーネント
 を、 $\Sigma_k(s)(i, i)$ は、分散マトリクス $\Sigma_k(s)$ の、第 i 行第 i 列のコン
 ポーネントを、それぞれ表す。そして、 k クラスモデルの出力確率は、これらに
 よって規定される。

【0040】

なお、HMMは、上述したように、初期状態確率 $\pi_k(q_h)$ 、遷移確率 $a_k(q_i, q_j)$ 、および出力確率 $b_k(q_i)$ (O) によって規定されるが、これらは、学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求めることとする。

【0041】

ここで、HMMとして、図5に示したものをを用いる場合には、常に、最も左の状態 q_1 から遷移が始まるので、状態 q_1 に対応する初期状態確率だけが1とされ、他の状態に対応する初期状態確率はすべて0とされる。また、出力確率は、式(9)、(10)から明らかなように、 $\Psi(t)(i, i)$ を0とすると、特徴ベクトルの分散を考慮しない場合の連続HMMにおける出力確率に一致する。

【0042】

なお、HMMの学習方法としては、例えば、Baum-Welchの再推定法などが知られている。

【0043】

図4に戻る。識別関数演算部21-k ($k=1, 2, \dots, K$) は、kクラスモデルについて、あらかじめ学習により求められている初期状態確率 $\pi_k(q_h)$ 、遷移確率 $a_k(q_i, q_j)$ 、および出力確率 $b_k(q_i)$ (O) によって規定される式(8)の識別関数 $G_k(Z)$ を記憶しており、特徴抽出部2からの特徴分布パラメータZを引数として、識別関数 $G_k(Z)$ を演算し、その関数値(上述した観測確率) $G_k(Z)$ を、決定部22に出力する。識別関数演算部21-sは、無音音響モデル補正部7から供給される初期状態確率 $\pi_s(q_h)$ 、遷移確率 $a_s(q_i, q_j)$ 、および出力確率 $b_s(q_i)$ (O) によって規定される、式(8)の識別関数 $G_k(Z)$ と同様の識別関数 $G_s(Z)$ を記憶しており、特徴抽出部2からの特徴分布パラメータZを引数として、識別関数 $G_s(Z)$ を演算し、その関数値(上述した観測確率) $G_s(Z)$ を、決定部22に出力する。

【0044】

決定部22では、識別関数演算部21-1乃至21-k、および識別関数演算部21-sそれぞれからの関数値 $G_k(Z)$ (ここでは、関数値 $G_s(Z)$ を含む

ものとする) に対して、例えば、次式 (11) に示す決定規則を用いて、特徴分布パラメータ Z 、すなわち、入力された音声に属するクラス (音響モデル) が識別される。

【数 11】

$$C(Z)=C_k, \text{ if } G_k(Z)=\max_i \{G_i(Z)\} \quad \dots (11)$$

ただし、 $C(Z)$ は、特徴分布パラメータ Z が属するクラスを識別する識別操作 (処理) を行う関数を表す。また、式 (11) の第 2 式の右辺における \max は、それに続く関数値 $G_i(Z)$ (ただし、ここでは、 $i = s, 1, 2, \dots, K$) の最大値を表す。

【0045】

決定部 22 は、式 (11) にしたがって、クラスを決定すると、それを、入力された音声の認識結果として出力する。

【0046】

図 1 に戻る。無音音響モデル補正部 7 は、ノイズ観測区間抽出部 3 から入力されるノイズ観測区間 T_n の音声データ (環境ノイズ) に基づいて、音声認識部 6 に記憶されている無音音響モデルに対応する識別関数 $G_s(Z)$ を生成して音声認識部 6 に供給する。

【0047】

具体的には、無音音響モデル補正部 7 では、ノイズ観測区間抽出部 3 から入力されるノイズ観測区間 T_n の音声データ (環境ノイズ) の M 個のフレームの各フレームについて、特徴ベクトル X が観測され、それらの特徴分布が生成される。

【数 12】

$$\{F_1(X), F_2(X), \dots, F_M(X)\} \quad \dots (12)$$

なお、特徴分布 $\{F_i(X), i = 1, 2, \dots, M\}$ は、確率密度関数 (Probabilistic Density Function) であるので、以下、無音特徴分布 PDF とも記述する。

【0048】

次に、無音特徴分布PDFを、次式(13)に従い、図7に示すように、無音音響モデルに対応する確率分布 $F_s(X)$ に写像する。

【数13】

$$F_s(X) = V(F_1(X), F_2(X), \dots, F_M(X)) \quad \dots (13)$$

ただし、 V は無音特徴分布PDF $\{F_i(X), i = 1, 2, \dots, M\}$ を無音音響モデル $F_s(X)$ に写像する補正関数(写像関数)である。

【0049】

この写像は、無音特徴分布PDFの記述によって様々な方法が考えられる。例えば、

【数14】

$$F_s(X) = \sum_{i=1}^M \beta_i(F_1(X), F_2(X), \dots, F_M(X), M) \cdot F_i(X) \quad \dots (14)$$

$$= \sum_{i=1}^M \beta_i \cdot F_i(X) \quad \dots (15)$$

ただし、 $\beta_i(F_1(X), F_2(X), \dots, F_M(X), M)$ は、各無音特徴分布の重み関数であり、以下、 β_i と記述する。なお、重み関数 β_i は、次式(16)の条件を満足するものである。

【数15】

$$\sum_{i=1}^M \beta_i(F_1(X), F_2(X), \dots, F_M(X), M) = \sum_{i=1}^M \beta_i = 1 \quad \dots (16)$$

【0050】

ここで、無音音響モデルの確率分布 $F_s(X)$ が正規分布であると仮定し、また、各フレームの特徴ベクトルを構成するコンポーネントが無相関であると仮定すれば、無音特徴分布PDF $\{F_i(X), i = 1, 2, \dots, M\}$ の共分散行列 Σ_i は対角線行列となる。ただし、この仮定的前提条件は、無音音響モデルの共分散行列も対角線行列であることである。したがって、各フレームの特徴ベクト

ルを構成するコンポーネントが無相関であれば、無音特徴分布PDF $\{F_i(X), i = 1, 2, \dots, M\}$ は、各コンポーネントに対応する平均と分散を持つ正規分布 $G(E_i, \Sigma_i)$ となる。 E_i は $F_i(X)$ の期待値であり、 Σ_i は $F_i(X)$ の共分散行列である。

【0051】

さらに、ノイズ観測区間 T_n の M 個のフレームに対応する無音特徴分布の平均を μ_i 、分散を σ_i^2 と表すことにすれば、無音特徴分布の確率密度関数は、正規分布 $G(\mu_i, \sigma_i^2)$ ($i = 1, 2, \dots, M$) と表すことができる。したがって、各フレームに対応する平均 μ_i 、および分散 σ_i^2 を用い、以下に示す様々な方法によって演算される無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ (上述した $G_s(Z)$ に相当する) は、図7に示した無音音響モデル $F_s(X)$ の近似分布となる。

【0052】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第1の方法は、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i = 1, 2, \dots, M\}$ を用い、次式(17)に示すように、全ての μ_i の平均を無音音響モデルの平均値 μ_s とし、次式(18)に示すように、全ての σ_i^2 の平均を無音音響モデルの分散 σ_s^2 とする方法である。

【数16】

$$\mu_s = \frac{a}{M} \sum_{i=1}^M \mu_i \quad \dots (17)$$

$$\sigma_s^2 = \frac{b}{M} \sum_{i=1}^M \sigma_i^2 \quad \dots (18)$$

ここで、 a および b は、シミュレーションにより最適な値が決定される係数である。

【0053】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第2の方法は、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i = 1, 2, \dots, M\}$ の期待値 μ_i のだけを用い

、次式(19)、(20)に従って、無音音響モデルの平均値 μ_s と、分散 σ_i^2 を演算する方法である。

【数17】

$$\mu_s = \frac{a}{M} \cdot \sum_{i=1}^M \mu_i \quad \dots (19)$$

$$\sigma_s^2 = b \cdot \left(\frac{1}{M} \cdot \sum_{i=1}^M \mu_i^2 - \mu_s^2 \right) \quad \dots (20)$$

ここで、aおよびbは、シミュレーションにより最適な値が決定される係数である。

【0054】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第3の方法は、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i=1, 2, \dots, M\}$ の組み合わせによって、無音音響モデルの平均値 μ_s と、分散 σ_s^2 を演算する方法である。

【0055】

この方法においては、各無音特徴分布 $G(\mu_i, \sigma_i^2)$ の確率統計量を X_i とする。

【数18】

$$\{X_1, X_2, \dots, X_M\} \quad \dots (21)$$

【0056】

ここで、無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ の確率統計量を X_s とすれば、確率統計量 X_s は、次式(22)に示すように、確率統計量 X_i と重み関数 β_i の線形結合で表すことができる。なお、重み関数 β_i は式(16)の条件を満足している。

【数19】

$$X_s = \sum_{i=1}^M \beta_i \cdot X_i \quad \dots (22)$$

【0057】

そして、無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ は、次式(23)に示すように表される。

【数20】

$$G(\mu_s, \sigma_s^2) = G\left(\sum_{i=1}^M \beta_i \mu_i, \sum_{i=1}^M \beta_i^2 \sigma_i^2\right) \quad \dots (23)$$

【0058】

なお、式(23)を一般化するためには、重み関数 β_i が、 $\{\beta_i = 1/M, i = 1, 2, \dots, M\}$ と仮定され、平均値 μ_s と、分散 σ_s^2 には、係数が乗算される。

【数21】

$$\mu_s = \frac{a}{M} \cdot \sum_{i=1}^M \mu_i \quad \dots (24)$$

$$\sigma_s^2 = \frac{b}{M^2} \cdot \sum_{i=1}^M \sigma_i^2 \quad \dots (25)$$

ここで、 a および b は、シミュレーションにより最適な値が決定される係数である。

【0059】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第4の方法では、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i = 1, 2, \dots, M\}$ の確率統計量 X_i に対応する統計母集団 $\Omega_i = \{f_{i,j}\}$ を仮定する。ここで、

【数22】

$$\{N_i \equiv N; i = 1, 2, \dots, M\}$$

とすれば、平均値 μ_i は、次式(26)によって得ることができ、分散 σ_i^2 は、次式(28)によって得ることができる。

【数 2 3】

$$\mu_i = \frac{1}{N} \sum_{j=1}^M f_{i,j} \quad \dots (26)$$

$$\sigma_i^2 = \frac{1}{N} \sum_{j=1}^M (f_{i,j}^2 - \mu_j^2) \quad \dots (27)$$

$$= \frac{1}{N} \sum_{j=1}^M f_{i,j}^2 - \mu_j^2 \quad \dots (28)$$

【0060】

式 (28) を変形すれば、次式 (29) の関係が成立する。

【数 2 4】

$$\frac{1}{N} \sum_{j=1}^M f_{i,j}^2 = \sigma_i^2 + \mu_i^2 \quad \dots (29)$$

【0061】

ここで、統計母集団の和 Ω

【数 2 5】

$$\Omega = \bigcup_{i=1}^M \Omega_i$$

を考慮すれば、式 (26) から次式 (30), (31) が導かれ、式 (29) から次式 (32) 乃至 (31) が導かれる。

【数 2 6】

$$\mu_s = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N f_{i,j} \quad \dots (30)$$

$$= \frac{1}{M} \sum_{i=1}^M \mu_i \quad \dots (31)$$

$$\sigma_s^2 = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f_{i,j} - \mu_s)^2 \quad \dots (32)$$

$$= \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N f_{i,j}^2 - \mu_s^2 \quad \dots (33)$$

$$= \frac{1}{M} \sum_{i=1}^M (\sigma_i^2 + \mu_i^2) - \mu_s^2 \quad \dots (34)$$

【0 0 6 2】

なお、実際には、式 (3 1) と式 (3 4) には係数が乗算されて用いられる。

【数 2 7】

$$\mu_s = \frac{a}{M} \sum_{i=1}^M \mu_i \quad \dots (35)$$

$$\sigma_s^2 = b \cdot \left(\frac{1}{M} \sum_{i=1}^M (\sigma_i^2 + \mu_i^2) - \mu_s^2 \right) \quad \dots (36)$$

ここで、a および b は、シミュレーションにより最適な値が決定される係数である。

【0 0 6 3】

また、次式 (2 7) に示すように、分散 σ_s^2 に対してだけ、係数を乗算するのうにしてもよい。

【数 2 8】

$$\sigma_s^2 = \frac{b}{M} \sum_{i=1}^M \sigma_i^2 + \frac{1}{M} \sum_{i=1}^M \mu_i^2 - \mu_s^2 \quad \dots (37)$$

【 0 0 6 4 】

次に、音声認識装置の動作について説明する。

【 0 0 6 5 】

フレーム化部 2 には、マイクロフォン 1 で集音された音声データ（環境ノイズを含む認識対象の発話音声）が入力され、そこでは、音声データがフレーム化され、各フレームの音声データは、観測ベクトル a として、ノイズ観測区間抽出部 3、および特徴抽出部 5 に順次供給される。ノイズ観測区間抽出部 3 では、発話スイッチ 4 がオンとされたタイミング t_b 以前のノイズ観測区間 T_n の音声データ（環境ノイズ）が抽出されて特徴抽出部 5 および無音音響モデル補正部 7 に供給される。

【 0 0 6 6 】

特徴抽出部 5 では、フレーム化部 2 からの観測ベクトル a としての音声データが音響分析され、その特徴ベクトル y が求められる。さらに、特徴抽出部 5 では、求められた特徴ベクトル y に基づいて、特徴ベクトル空間における分布を表す特徴分布パラメータ Z が算出され、音声認識部 6 に供給される。音声認識部 6 では、特徴抽出部 5 からの特徴分布パラメータを用いて、無音区間および所定数 K の単語それぞれに対応する音響モデルの識別関数の値が演算され、その関数値が最大である音響モデルが、音声の認識結果として出力される。また、音声認識部 6 では、無音音響モデル補正部 7 から入力される無音音響モデルに対応する識別関数を用いて、それまで記憶されていた無音音響モデルに対応する識別関数が更新される。

【 0 0 6 7 】

以上のように、観測ベクトル a としての音声データが、その特徴量の空間である特徴ベクトル空間における分布を表す特徴分布パラメータ Z に変換されるので、その特徴分布パラメータは、音声データに含まれるノイズの分布特性を考慮したものとなっており、また、無音区間を識別するための無音音響モデルに対応する識別関数が、発話直前のノイズ観測区間 T_n の音声データに基づいて更新されているので、音声認識率を大きく向上させることが可能となる。

【0068】

次に、図8は、発話スイッチ4がオンとされてから発話が始まるまでの無音区間 T_s を変化させたときの音声認識率の変化を測定した実験の結果を示している。

【0069】

なお、図8において、曲線aは無音音響モデルを補正しない従来の方法による結果を示しており、曲線bは第1の方法による結果を示しており、曲線cは第2の方法による結果を示しており、曲線dは第3の方法による結果を示しており、曲線eは、第4の方法による結果を示している。

【0070】

実験の条件は、以下の通りである。認識される音声データは、高速道路を走行中の車内で集音されたものである。ノイズ観測区間 T_n は、20フレームで約0.2秒である。無音区間 T_s は、0.05秒、0.1秒、0.2秒、0.3秒、0.5秒とした。音声データの特徴抽出においては、MFCC(Mel-Frequency Cepstral Coefficients)ドメインで分析を実施した。認識の対象とする音声の発話者は、男女4人ずつ計8人であり、一人当たり303個の単語を離散して発話した。タスクは大語彙離散日本語で5000ワードである。音響モデルは、HMMであり、良好な音声データを用いて予め学習されているものである。音声認識においては、Viterbiサーチ法でビーム幅を3000とした。

【0071】

なお、第1、第2、および第4の方法においては、係数 a を1.0とし、係数 b を0.1とした。第3の方法においては、係数 a を1.0とし、係数 b を1.0とした。

【0072】

図8から明らかなように、従来の方法(曲線a)では、無音区間 T_s が長くなるのに伴って音声認識率が著しく低下しているが、本発明の第1乃至4の方法(曲線b乃至e)では、無音区間 T_s が長くなっても、音声認識率は、わずかに低下しない。すなわち、本発明によれば、無音区間 T_s が変化しても、音声認識率はある程度のレベルを維持することが可能である。

【 0 0 7 3 】

以上、本発明を適用した音声認識装置について説明したが、このような音声認識装置は、例えば、音声入力可能なカーナビゲーション装置、その他各種の装置に適用可能である。

【 0 0 7 4 】

なお、本実施の形態では、ノイズの分布特性を考慮した特徴分布パラメータを求めるようにしたが、このノイズには、例えば、発話を行う環境下における外部からのノイズの他、例えば、電話回線その他の通信回線を介して送信されてくる音声の認識を行う場合には、その通信回線の特性なども含まれる。

【 0 0 7 5 】

また、本発明は、音声認識の他、画像認識その他のパターン認識を行う場合にも適用可能である。

【 0 0 7 6 】

なお、上記各処理を行うコンピュータプログラムは、磁気ディスク、CD-ROM等の情報記録媒体よりなる提供媒体のほか、インターネット、デジタル衛星などのネットワーク提供媒体を介してユーザに提供することができる。

【 0 0 7 7 】

【発明の効果】

以上のように、請求項 1 に記載のパターン認識装置、請求項 6 に記載のパターン認識方法、および請求項 7 に記載の提供媒体によれば、データが入力される直前に入力されたノイズに基づいて、データが存在しない状態に対応するモデルを生成し、記憶されている対応するものを更新するようにしたので、音声認識が開始された時から発話が始まるまでの時間が長くなるに伴って認識性能が低下することを抑止することが可能となる。

【図面の簡単な説明】

【図 1】

本発明を適用した音声認識装置の構成例を示すブロック図である。

【図 2】

図 1 のノイズ観測区間抽出部 3 の動作を説明するための図である。

【図 3】

図 1 の特徴抽出部 5 の詳細な構成例を示すブロック図である。

【図 4】

図 1 の音声認識部 6 の詳細な構成例を示すブロック図である。

【図 5】

音声認識部 6 の動作を説明するための図である。

【図 6】

図 1 の無音音響モデル補正部 7 の動作を説明するための図である。

【図 7】

図 1 の無音音響モデル補正部 7 の動作を説明するための図である。

【図 8】

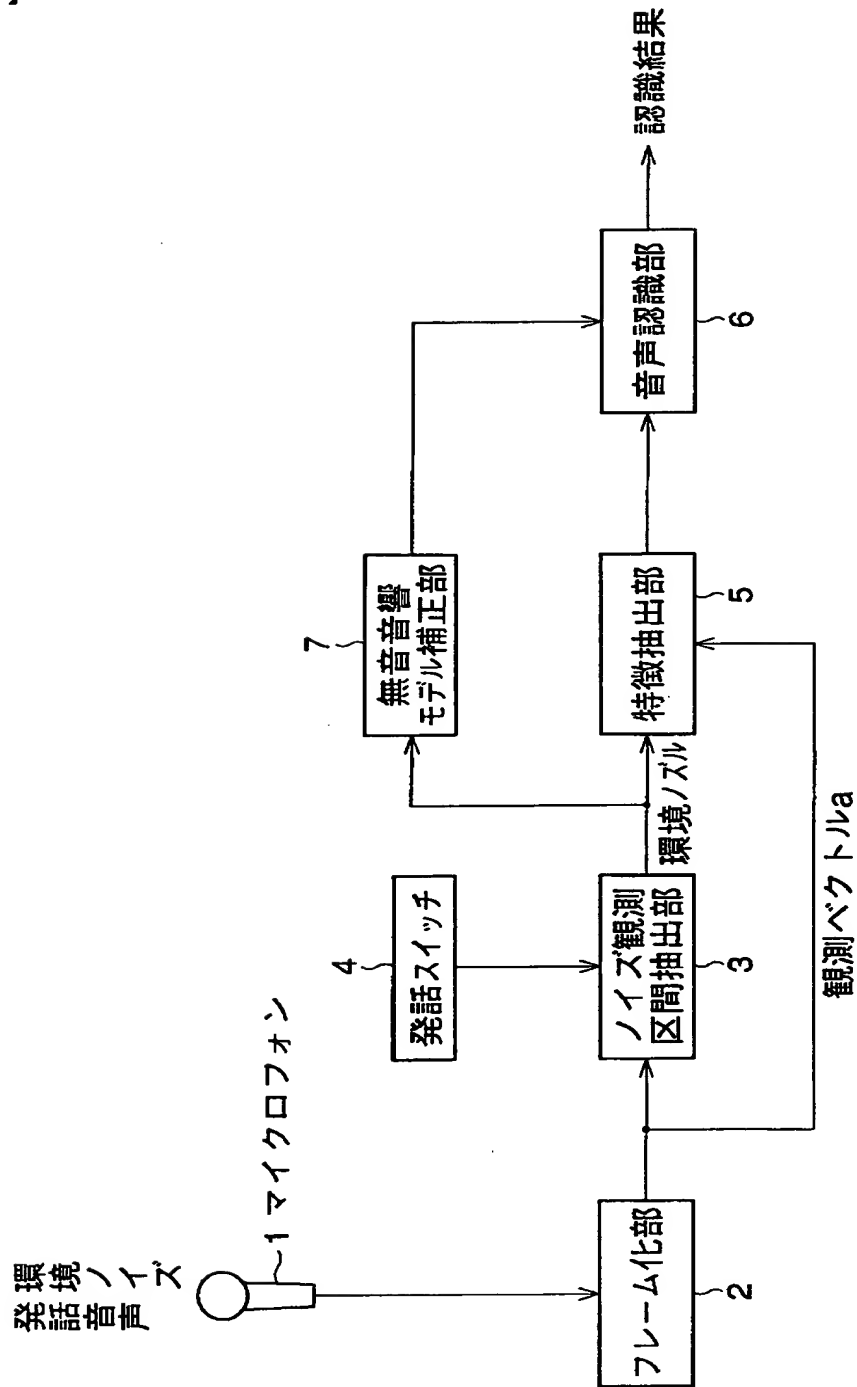
本発明を適用した音声認識装置の音声認識実験の結果を示す図である。

【符号の説明】

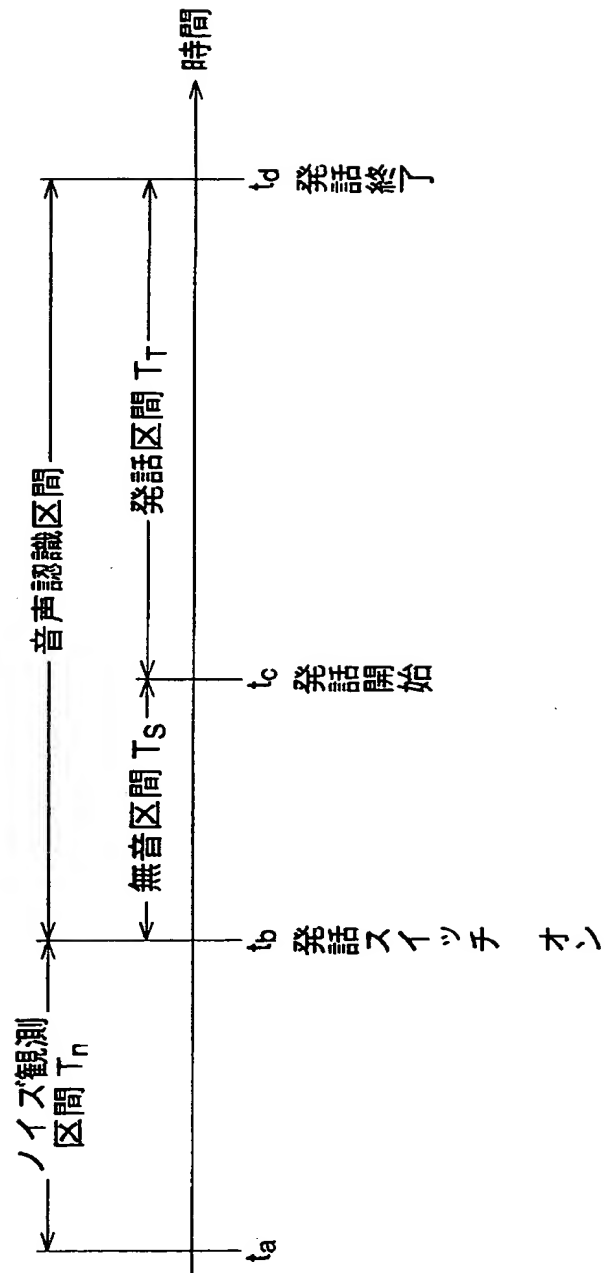
1 マイクロフォン, 2 フレーム化部, 3 ノイズ観測区間抽出部,
4 発話スイッチ, 5 特徴抽出部, 6 音声認識部, 7 無音音響モデル補正部

【書類名】図面

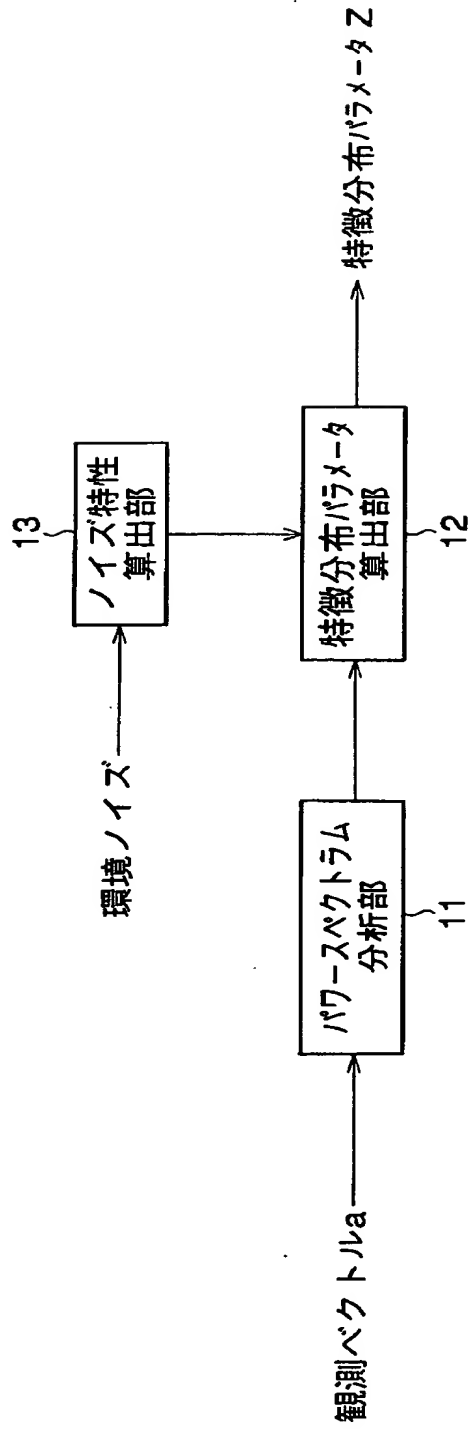
【図 1】



【図 2】

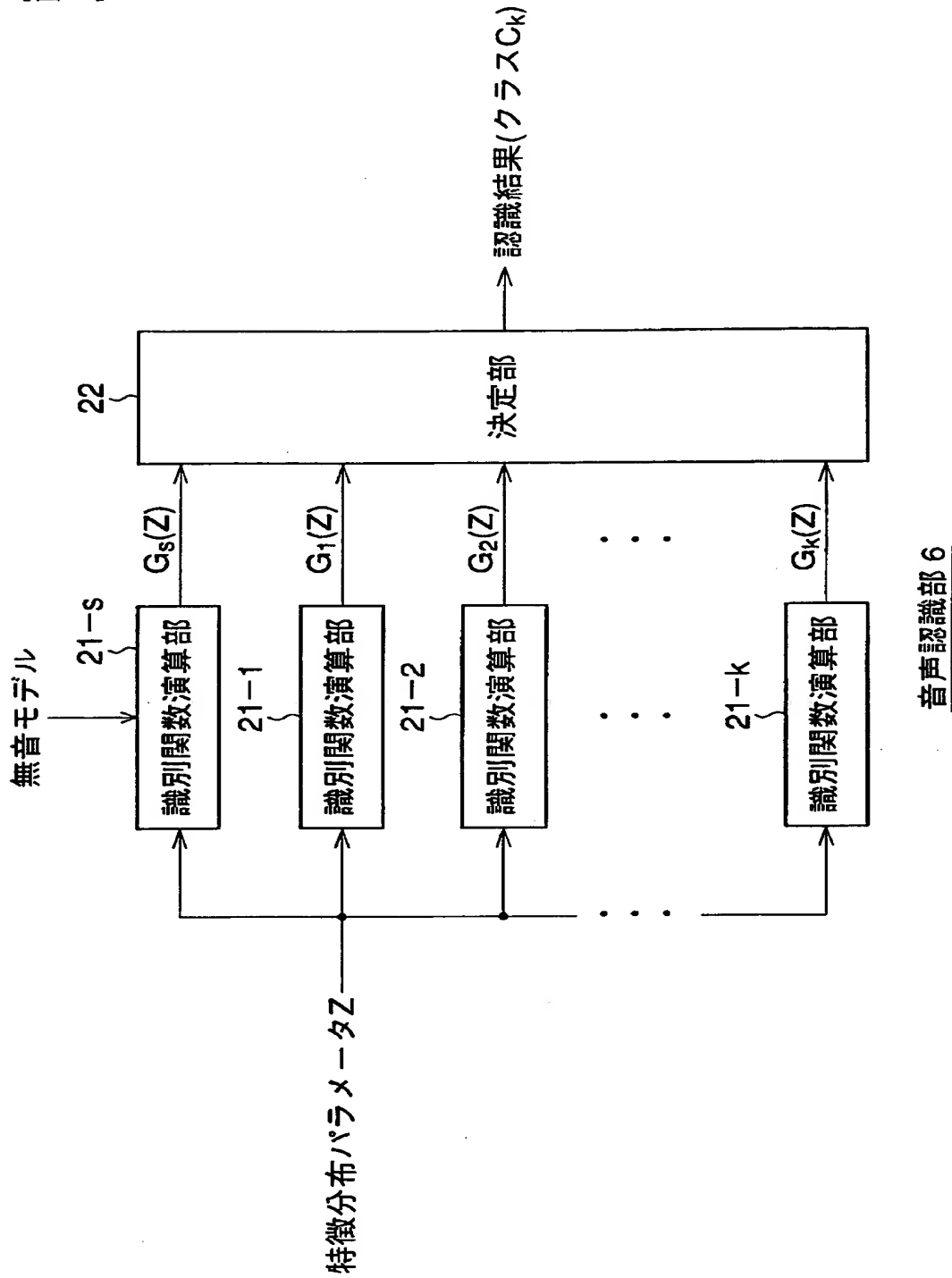


【図 3】

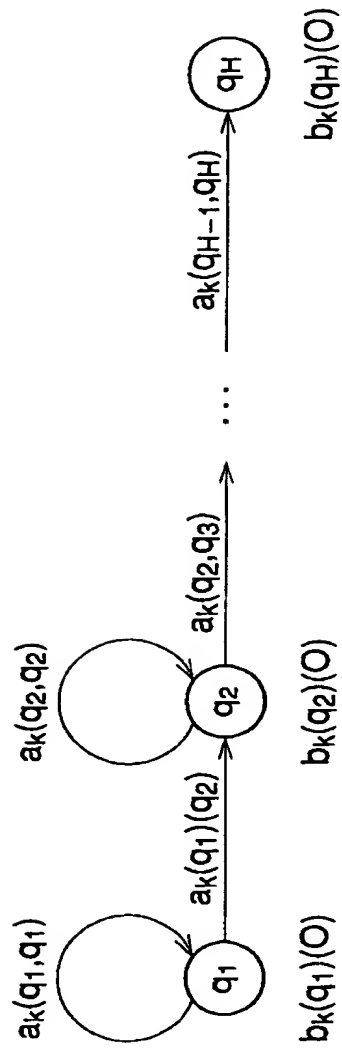


特徴抽出部 5

【図 4】

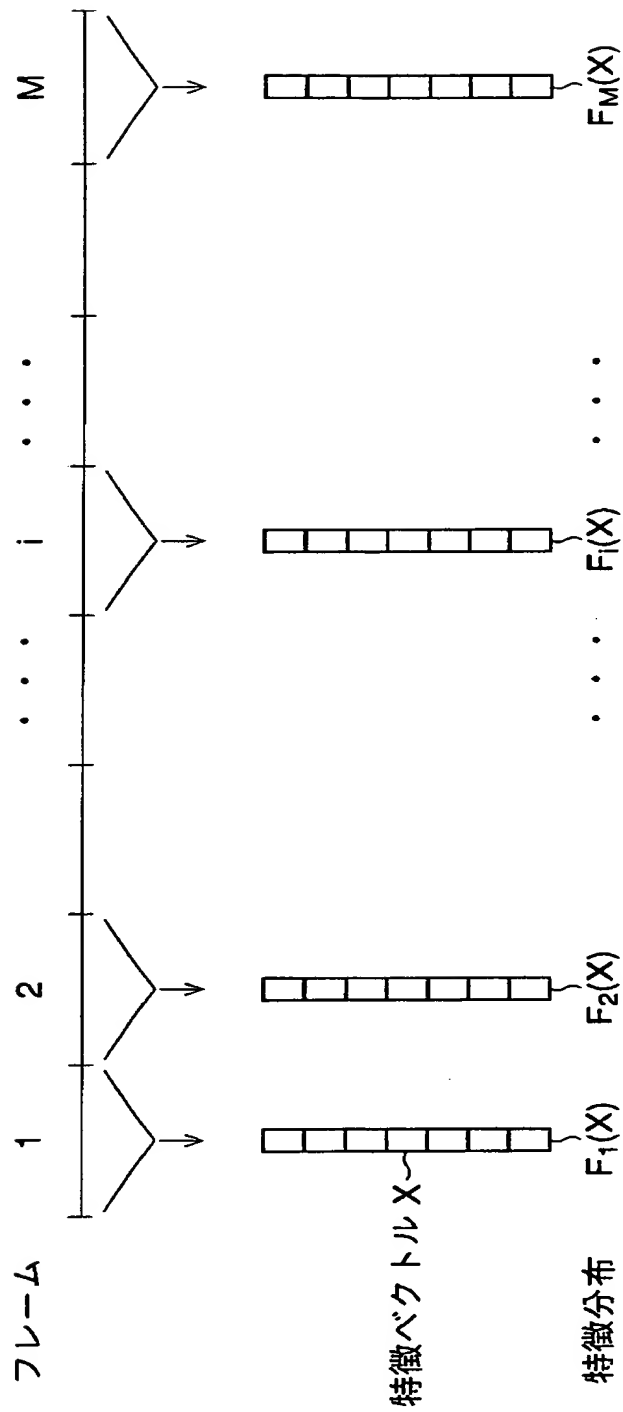


【図 5】

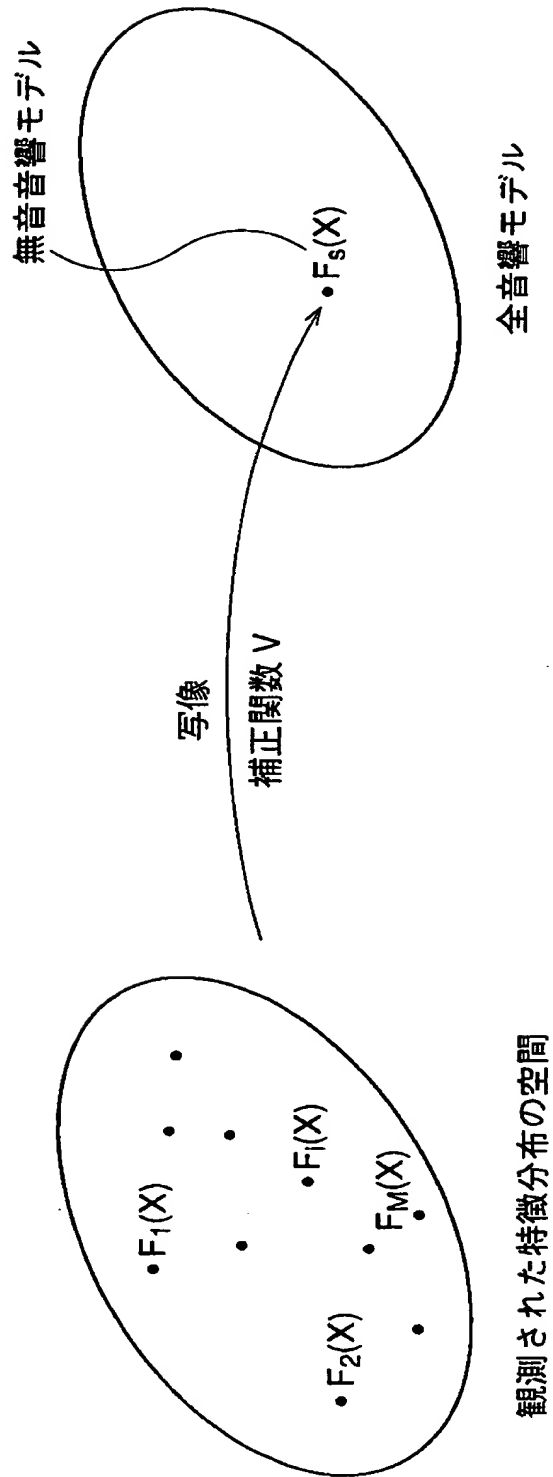


HMM(left-to-right モデル)

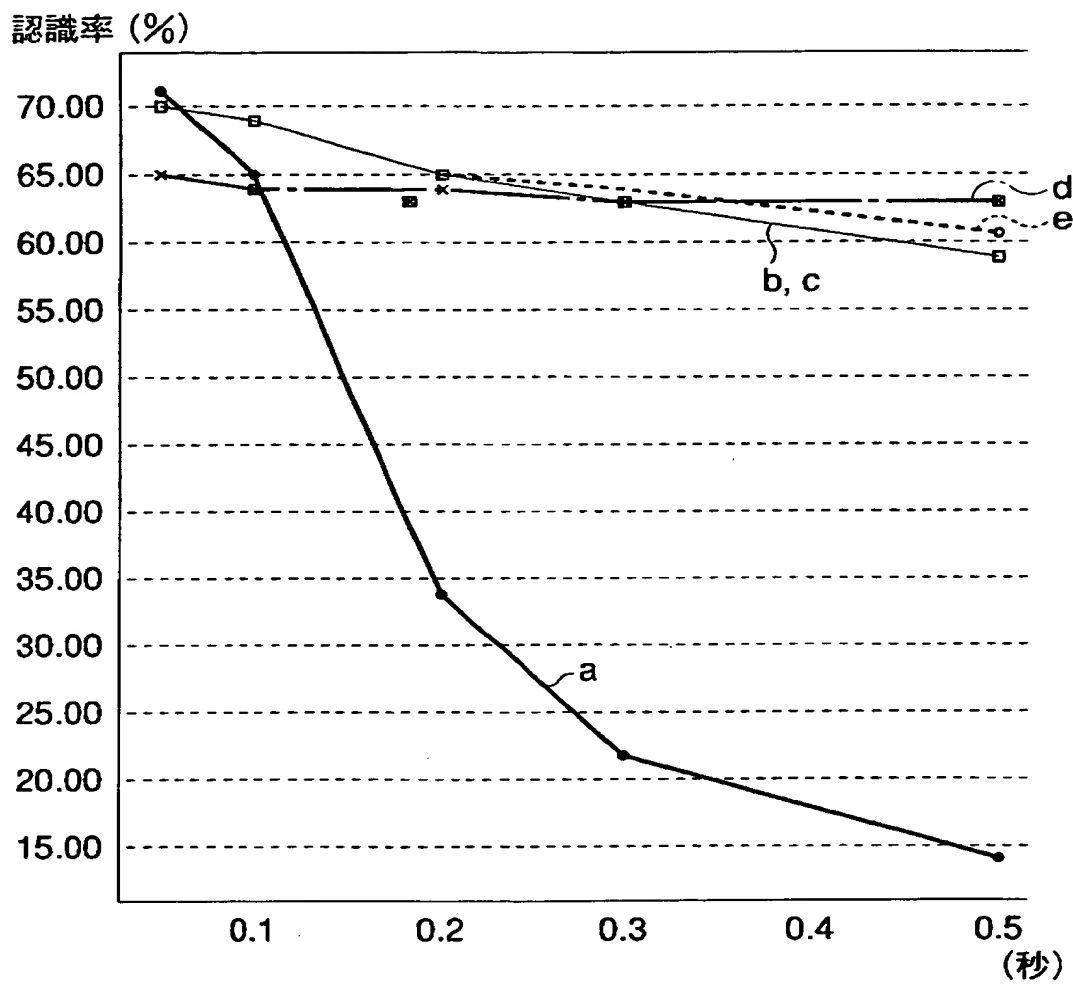
【図 6】



【図 7】



【図 8】



【書類名】 要約書

【要約】

【課題】 音声認識が開始された時から発話が始まるまでの時間が長くなるに伴って認識性能が低下することを抑止する。

【解決手段】 無音音響モデル補正部は、音声認識区間の初期に存在する無音区間 T_s を認識するために用いられる無音音響モデルを、ノイズ観測区間 T_n に含まれる環境ノイズに基づいて生成する。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号

[000002185]

1. 変更年月日	1990年 8月30日
[変更理由]	新規登録
住 所	東京都品川区北品川6丁目7番35号
氏 名	ソニー株式会社